

# A New Source-Splitting Approach to the Slepian-Wolf Problem

Todd Coleman, Anna Lee, Muriel Médard; Michelle Effros  
 Lab for Information and Decision Systems; Data Compression Laboratory  
 Massachusetts Institute of Technology; California Institute of Technology  
 Cambridge, MA 02139; Pasadena, CA 91125  
 {colemant,annalee,medard}@mit.edu;effros@caltech.edu

**Abstract** — It is shown that achieving an arbitrary rate-point in the achievable region of the  $M$ -source Slepian-Wolf [1] problem may be reduced via a practical *source-splitting* transformation to achieving a corner point in a  $2M - 1$  source Slepian-Wolf problem. Moreover, each source must be split at most once. This approach extends the ideas introduced in [2] to a practical setting: it does not require common randomness shared between splitters and the decoders, the cardinality of each source split is strictly smaller than the original, and practical iterative decoding methods can achieve rates near the theoretical bound.

## I. INTRODUCTION

We introduce a new source-splitting technique to transform achieving any rate pair in the  $M$ -source Slepian-Wolf to achieving a corner point in a  $2M - 1$  source Slepian-Wolf problem. Corner points are attractive because of the ability to pipeline the decoding process. This technique is a practical manifestation of the ideas introduced in [2]: it does not require the sharing of common randomness between encoders and the decoder, the cardinalities of each split are smaller than the original, and pipelinable low-complexity iterative decoding techniques can be applied to achieve rates near the theoretical bound.

Another corollary of the cardinality reduction of this approach is that decoding for the arbitrary-source, arbitrary alphabet Slepian-Wolf problem is only as complex as binary source-coding with side information for a single source.

## II. PROPOSED SOURCE-SPLITTING APPROACH

We now consider a discrete memoryless source (DMS)  $U = (U_1, U_2, \dots)$  where  $U_i \in \mathcal{Q} = \{0, 1, \dots, Q - 1\}$ . We say that  $U_i \leftrightarrow (U_i^a, U_i^b)$  if there is a bijection between  $U_i$  and  $(U_i^a, U_i^b)$ . Such a split can be constructed as follows:

$$U_i \mapsto \begin{pmatrix} U_i^a = \min(\pi(U_i), T) \\ U_i^b = \max(\pi(U_i), T) - T \end{pmatrix} \quad (1)$$

$$\mapsto U_i = \pi^{-1}(U_i^a + U_i^b) \quad (2)$$

where  $T \in \mathcal{Q}$  and  $\pi$  is a permutation of  $\mathcal{Q}$ . We note that (1) provides  $2^Q - Q - 2$  distinct splits.

## III. SPLITTING TO YIELD $2M - 1$ SOURCE CORNER POINTS WITH AT MOST ONE SPLIT PER SOURCE

Now we show that by performing the splitting operation across  $n$  outcomes of the source  $U$ , the average entropy of  $U^a$  may lie arbitrarily close to any point in  $[0, H(U)]$ . Let us consider  $\underline{U}_{[n]} = (U_1, \dots, U_n)$  as the single outcome of a DMS with alphabet  $\{0, \dots, Q^n - 1\}$ . We note that there are approximately  $2^{nH(U)}$  ‘typical’ sequence outcomes and

$$Q^n - 2^{nH(U)} = 2^{nH(U)} \left( 2^{n[\log_2 Q - H(U)]} - 1 \right)$$

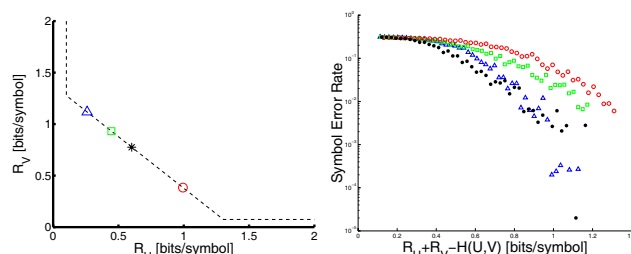


Figure 1: Symbol error rate for source-splitting to achieve non-corner rate pairs.

‘atypical’ ones. For a given  $\epsilon \in [0, 1]$  we select  $\pi$  and  $T$  so that  $2^{n\epsilon H(U)}$  typical and  $2^{n\epsilon H(U)} \left( 2^{n[\log_2 Q - H(U)]} - 1 \right)$  atypical sequences take values below  $T$ . We see that by varying  $\epsilon, \pi, T$  the closure of all average entropies  $\frac{1}{n} H(\underline{U}_{[n]}^a)$  is  $[0, H(U)]$ .

For the two-source Slepian-Wolf problem, sources  $U$  and  $V$  distributed according to  $P(U, V)$  are separately encoded at achievable rates  $(R_1, R_2)$ . Any point on the dominant face may be achieved by splitting  $U \mapsto (U^a, U^b)$  according to (1), assigning encoders according to

$$R_1^a = H(U^a), R_2 = H(V|U^a), R_1^b = H(U^b|U^a, V)$$

where  $R_1 = R_1^a + R_1^b$ . We note that  $(R_1^a, R_2, R_1^b)$  is a corner point for the 3-source Slepian-Wolf problem of encoding  $(U^a, V, U^b)$  and thus decoding may be done in a low-complexity pipelined fashion exploiting the chain rule for entropy. By varying  $\epsilon, \pi, T$  as above, we may achieve any rate pair  $(R_1, R_2)$  on the dominant face of the Slepian-Wolf region. Figure 1 gives simulation results for applying the pipelined decoding using the sum-product algorithm for LDPC syndrome-former encoding and illustrates the practicality of this approach. The general  $M$ -source case follows as a direct extension of the preceding arguments.

Any achievable rate can be achieved by splitting each source at most once. Algebraic topology techniques used to prove the analogous result in the discrete multiple access setting ([3], sec. III) directly apply in this setting.

## REFERENCES

- [1] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471-480, 1973.
- [2] B. Rimoldi and R. Urbanke. Asynchronous Slepian-Wolf coding via source-splitting. *ISIT*, Ulm, Germany, June 29-July 4 1997, p. 271.
- [3] A. Grand, B. Rimoldi, R. Urbanke, and P. A. Whiting. Rate-splitting multiple access for discrete memoryless channels. *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 873-890, 2001.